Annamalai University Publications in Linguistics No. 6

6300

LEXICOGRAPHY

MAHAMAHOPADHYAYA
Dr. U. V. SWAMINATHAIYAR LIBRARY.
TIRUVANMIYUR, MADRAS-41.

22942

BY

S. M. KATRE

P31, DX1
K5

ANNAMALAI UNIVERSITY
ANNAMALAINAGAR
1965

LEXICOGRAPHY

(A course of three special lectures delivered at the Linguistics Department of the Annamalai University)

BY

SUMITRA MANGESH KATRE,

M. A., Ph. D. (London)

Director & Honorary Professor of Linguistics

Director, Centre of Advanced Study in Linguistics, Deccan College of Post - Graduate & Research Institute, Poona.

Editor, Dictionary of Sanskrit on Historical Principles.

ANNAMALAI UNIVERSITY

ANNAMALAINAGAR (MADRAS STATE)

1965

₨ 4 - 00

Githa Press, Tiruchi-2

PUBLICATIONS IN LINGUISTICS NO. 6.

PREFACE

Dr. S. M. Katre who is responsible to a great extent for creating a new interest in Linguistics after Independance was also partly responsible for organising the Linguistics Department at the Annamalai University. He is the editor of the great Sanskrit Dictionary which is now in progess as a national venture. The Linguistics Department of the Annamalai University was then thinking of a project of a Tamil Lexicon on Historical principles. Therefore Dr. S. M. Katre was requested to deliver three extension lectures on Lexicography. In modern India there are various projects for preparing dictionaries in the different languages. Therefore these lectures which are the results of his practical experience, will be useful for all those who are interested in Lexicography. I am happy that these lectures are now made available in the printed form.

Annamalainagar
16—12—1965.

T. P. MEENAKSHISUNDARAM.

LEXICOGRAPHY

I

1.1. A lexicographer, according to the colourful definition of Dr. Samuel Johnson, is a harmless drudge. While it is possible to differ from that distinguished Doctor on the term 'harmless', those especially concerned with lexicographical activities will agree whole-heartedly with the second term. Lexicography is an art which has been developed over a fairly long period of human endeavour, and in a sense precedes man's activities in the field of linguistics. We have evidence of this in ancient Sanskrit and Greek literatures. Here in India our tradition has been principally oral, but nevertheless a sizeable body of literature grew from the very beginning, transmitted through a never-ending teacher-pupil relationship. We have a story narrated by Patañjali in the Mahābhāṣya that Bṛhaspati started teaching Indra, king of the Gods, the recitation of individual words found in Sanskrit; the period of study extended over a thousand divine years, but Indra did not come to the end of his task! A great deal of truth is hidden in this statement. In the first place the study of individual words in a language preceded an analysis of the forms to be found in it. The interest of the speaker in his language stemmed from the fact that when he inherited a literature, whether oral or written, which exhibited forms and expressions which were

no longer current, he was forced to enter more deeply into questions relating to the origin of such forms and expressions. This resulted in the development of a technique of analysis which has been called Vyākaraṇa in India and Grammar in other countries deriving their inspiration from Greece and Italy. Indra, once again, according to a story in the Brāhmaṇas, was requested by the gods to analyse language, and the first description of Sanskrit is ascribed to him. Indian tradition makes Indra the first analyser of language. Considering the fact that scientific linguistics was born in India more than thirty centuries back we may well deify the first scholar who discovered the analytic approach to linguistic studies.

1.2. The earliest form in which Lexicography developed was through making special note of remarkable forms and expressions which required an explanation or gloss to be understood by the speakers of the day, and these usually consisted of obsolete or unusual forms. Thus in India, in so far as Vedic literature was concerned, the Nighaṇṭus were compiled as an ancillary branch of Vedic tradition. These consisted of strings of words under certain main ideas: for instance there are 21 names for "earth" 101 for "water", 16 for "air" and 122 verbs for "to go". These texts were similarly transmitted by verbal tradition in regular Vedic schools, and became the basis of Yāska's Nirukta which is the only one of the Vedāṅga Nirukta that we possess. Here is another proof

that such collections preceded the actual analysis of language; etymology as an accessory to linguistic analysis developed after the beginning of lexicography had collected suitable material for such analysis. In ancient Greece and Rome where the manuscript transmission of texts was more vital it was not unusual for a learned scribe or copyist making a new copy of an exemplar before him of an early text to insert either in the margin or in between the lines themselves a simple gloss to explain a difficult or obsolete expression found in the text. Indeed, this practice of inserting marginal or interlineal glosses continued right up to the medieval times, principally in relation to Latin texts, but also in connection with the older forms of modern European languages. Modern scholars seem to have inherited this faculty from their medieval ancestors in all parts of the world; witness, for example, the language texts which are prescribed for various examinations from the school to the college level: it would be difficult to find an exception to this rule of marginal or interlinear insertions of meanings and explanations! When this marginal and interlinear glossing had become a fairly well-established practice, and libraries began to be built up, the next step appears to be the collection of such glosses from different texts and the first glossaries begin to appear. They are simple in structure and confine themselves only to giving single equivalents to each form or expression. Obviously in India as in ancient Greece, the explanations were in the newer form of the same

language: it was monolingual or unilingual, the obsolete or unusual form being replaced by the more current form. On the other hand when Sanskrit and Latin ceased to be vehicles of literary communication but became merely the learned medium of scholarly instruction, and newer literary languages occupied their place in society, the glosses appended to the texts were naturally in the classical languages. Thus, for instance, for Kannada we have the earliest grammatical description as well as lexicographical works through the medium of Sanskrit. Obviously, therefore, the glossing was not the work of a professional copyist but that of a professed scholar who inserted his gloss. When his exemplar became the source of a new manuscript transcription the gloss was either copied as a marginal or interlinear entry as in the original or became a part of the running text in the same line. Examples of this can be cited by every scholar who works with manuscripts of especially difficult texts,

1.3. In Europe it was principally the study of classical languages which necessitated the glossing process. This habit of glossing manuscripts of Latin writings such as the gospels and the psalter resulted ultimately in an interlinear translation of the original, giving a one to one correspondence between the classical language and the one used for translation. Soon such glosses were used by studious men as a means of increasing their classical vocabulary; for this purpose a man

would abstract the words from the volume and make a list of them with their meanings in the other language. In this process he might collate similar glosses from other texts, giving rise to a fairly scattered number of simple and complex lists. Occasionally the source of the text is indicated but this is by no means common practice. It took a number of generations to achieve this process before alphabetization of the complex lists began. Even this process of alphabetization involved a considerable time and passed through several stages. In the first process words beginning with A were separated from those beginning respectively with B, C and others and put in blocks without regard to the second letters of the words. Later on the second letters were alphabetized under each of the first letters, leaving the third and subsequent letters out of consideration. Finally alphabetization was carried out throughout as in a modern glossary. In this manner, for instance, in England by 1450 not only Latin-English lexicons were in use, but also English-Latin dictionaries were being made; and in the sixteenth century French-English dictionaries were being made on the same principles.

1.4. The lexicographical tradition in India closely parallels that of Vedic and Classical Sanskrit tradition. Being primarily an oral tradition, from the simple Nighaṇṭus to the classical lexicons, the form of the lexion changed in tune with the spirit of the times. The nature

of glossing may be seen in the development of the Brahmana literature in which we discover the first attempts at interpretation of earlier Vedic texts. Etymologies seek to find out the significance of verbal forms; from etymologies to a purposeful analysis of language led to the discovery of articulatory phonetics and a special ancillary branch of Vedic studies: 'Śikṣa and the Pratiśakhyas; and from phonetic analysis to a descriptive statement of the language itself, culminating in the system of Vyakaraṇa whose most significant representative is the system of Paṇini and the starting point in Modern Descriptive Linguistics. Pāṇini's analysis of Sanskrit not only indicates that it is the end result of a continuous process of analytical study but also attests to a collection of analysed material into systematic groups such as the Gaṇapāṭha and Dhātupāṭha. Indeed the generally accepted theory is that the Dhātupāṭha originally was merely a string of verbal bases classified and ranked according to the needs of grammatical statement, and it was only in the 3rd century *A. D.* that meanings were attached to the bases by Bhīmasena and these have since been transmitted with meanings.

The later lexicons in India are all set in verse form: they are either synonymous or homonymous. While the arrangement is by subjects in a majority of cases, some have attempted further classification by grouping words of one, two, three and more syllables; within each group, the further arrangement has been governed by the initial or

the final syllable, and some kind of alphabetisation has been attempted. A similar process has taken place in relation to Tamil Lexicograppy. Beginning with the Tolkāppiyam, in sūtra style in the early centuries of the Christian era, we have the Tivakaram & Pinkalantai as regular lexicons in sūtra form : later texts like Nikantu–cūṭāmaṇi are set in verse form. The Akarati-nikantu of Chidambara Revana Siddhar (*c.* 1594 *A.D*,) arranges its vocabulary according to the alphabetical arrangement only with regard to the initial syllables ; further alphabetization has not yet taken place. In the Potikainikantu of Swaminatha Kavirayar (1750 *A.D.*) alphabetisation has progressed up to the second syllable of each word. It was only when the European scholars came to study Indian languages and prepared suitable grammars, texts and glossaries or lexicons that the model developed in Europe was applied to Indian conditions and lexicons as works of reference with complete alphabetization came into being. It would thus appear that as long as transmission of texts was through oral tradition or the laborious process of hand-transcription, memorising took precedence over mere reference handbooks, and lexicons were compiled to satisfy the existing need.

1.5. With this brief indication of the background for lexicography we may now proceed to consider the place of this art and science in our modern context. While lexicography naturally preceded the regular analysis of language and in

fact led to it in its various aspects, its position in the discipline which constitutes Modern Linguistics has not been sufficiently realized, Lexicography is both a science and an art. However its place as a science may better be discussed under the more general term Lexicology, while as an art it may be considered under its general name of Lexicography. To the creative linguist lexicography is an essentially secondary discipline — the art or craft of manufacturing certain tools of handy information, while Lexicology is the science which directs the manner in which lexicography is to be attempted. It rises above the genre of individual languages and considers language as a single activity or mode of behaviour as a means of communication, while lexicography is only its application in a given instance. Lexicography has more often than not attracted slightly plodding or sometimes pedestrian workers, the number of specialists being few and far between.

1 6. Modern Linguistics has amply demonstrated that language is a complex system of habits which can be broken down into five principal sub-systems:

(1) The GRAMMATICAL sub-system which consists of a stock of morphemes and the arrangements in which they occur.

(2) The PHONOLOGICAL sub-system which consists of a stock of phonemes and the arrangements in which they occur.

(3) The MORPHOPOHNEMIC sub-system which is the code which ties together the GRAMMATICAL and the PHONOLOGICAL sub-systems.

(4) The SEMANTIC sub-system which associates various morphemes, combinations of morphemes and arrangements in which morphemes may be put, with things and situations, or kinds of things and situations.

(5) The PHONETIC sub-system which consists of the ways in which sequence of phonemes are converted into sound-waves by the articulation of a speaker and are decoded by a hearer from the speech signal.

Viewed from this angle the lexicographer has necessarily to deal with all the five sub-systems in order to achieve his objective. While these systems are strictly linguistic the first three sub-systems constitute the purely linguistic level while the last two cover levels which strictly belong to sister sciences but at the same time impinge on the first three which may, therefore, be called the central sub-systems of the speech world.

Language may be analysed at three levels. Since language may be defined as a system of arbitrary vocal symbols by means of which human beings communicate and cooperate with one another, we may analyse it at a given point of time in a given region : the result will be an analysis of the language which may be defined as

a synchronic description relating to the particular region. If the language has certain varying forms associated with distinct regions, a description may be attempted at a given point of time, covering the entire area consisting of the different regions, and we arrive at a synchronous linguistic geography of the language. If the language has been attested and preserved in literary documents spread over a range of time, we may analyse the language over its time scale and we get a diachronic description We have thus two scales with which we can operate : time and space (or region) and distinguish between the synchronic and the diachronic descriptions. The first may be called the static description while the second may be called the dynamic description of the language. A third level results when we compare the synchronic or diachronic descriptions of languages which exhibit a one-to-one correspondence in their central sub-systems and attempt a comparative statement at the phonological and morphological levels. This is particularly involved in the development of the so-called etymological lexicon which gives the prehistory of the vocables of a given language on the basis of such comparison within the family of which it is a particular realization. Thus we build up a descriptive, historical and comparative approach to the study of language with reference to each of its five sub-systems.

In relation to these approaches we have also a similar division with regard to Lexicography. Lexicography principally deals with the collection

of the significant forms or morphemes of the language and their ordering in the linguistic corpus with reference to the meaningful situations in which each occurs. In common parlance we say that a lexicon consists of a body of words or vocables (whence vocabulary) alphabetically arranged, with a list of meanings which they bear in significant situations; but the concept of the word itself is something which goes beyond the linguistic units specified in the sub-systems indicated above. The word is nothing more than a conventional or arbitrary segment of utterance; it may be briefly defined as a minimum free form, consisting of one or more morphemes. A morpheme is a minimum significant or meaningful unit, and it may be either BOUND or FREE. The object of the lexicon is then to serve as a repository of as many free forms as possible to be found in the language, and to assess its function in communication systems in relation to each form; indeed the lexicon cannot exclude bound forms, but must provide for an entry in each case to indicate its particular function. This may take the synchronic level, the diachronic level or the comparative etymological level. Lexicons can thus serve all the three levels of approach.

1.7. Given the synchronic approach we have two methods of operation. If we are dealing with a living situation within our own experience, as for instance, with a modern language, and the present time, we have two approaches possible. If the language is transmitted orally and does not

possess written records we operate at the idiolect and dialect level, gathering our material from actual speech events or utterances. If, on the other hand, it is transmitted both orally and through written symbols, written literature, we may either concentrate on the literature transmitted in writing or on actual utterances. The chances are, in the majority of cases exemplified in our country, in favour of concentration on the written records to the omission of spoken utterances or only partial inclusion of dialectal material. Consequently most of our existing lexicons in Indian languages with a literary background are primarily based on selective material from the literary standard, with only occasional inclusion of sub-standard and regional forms.

If a language has a long literary history behind it the lexicon will naturally operate on a selection of material from the available corpus of its literature, along with current literary standard forms. The lexicographer here operates on the diachronic scale. But in as much as the literature represents but one standardised source it may fail to illustrate dialectal variation.

Finally the comparative linguist draws his material from all available sources pertaining to the family or languages with which he operates in order to reach beyond the forms attested therein, and in doing so he is forced to reconstruct intermediate or ultimate forms to explain the individual realization

in each member as a result of unbroken verbal transmission along particular lines of tradition.

1·8. Let us assume that we as linguists are interested in compiling a lexicon of the particular language which we have subjected to a rigorous analysis, and let us further assume that this language is one which has not yet been reduced to a writing system. Our first task then is to devise such a writing system after making personal records of the utterances of a single speaker and subjecting them to continuous phonetic, phonemic and morphophonemic analysis of the records, checking them, if possible with another speaker. Once an analysis of the three central sub-systems has been made and the language given its characteristic writing system we begin to build up a corpus of texts from the utterance-events recorded by us. The larger this corpus the more inclusive will be our material for lexical purposes. The free forms can be extracted from this corpus and will serve as our lexical units or lexemes. Since each lexeme has a certain function or functions in the utterance-events the study of these function(s) will always accompany the extraction of the lexeme itself. If L represents a lexeme and F its functional meaning in a given situation or context, the purpose of the lexicon we wish to build up is to make as complete a repertory of the L's in relation to the F's, and the larger the corpus the larger will be the number of L's. In the case of a living and dynamically changing language it follows therefore that its lexicon is essentially unstable, preventing

establishment of any definitive repository of all its lexemes. But the success of the lexicon, if completeness is aimed at, will be in proportion ot the number of L's and in their significant situations each of which must find a place in it. Our chief witnesses in the case under consideration are the speakers of the language, and we check one witness with another to arrive at an acceptable statement about the linguistic character of the language.

1.9. In a situation where we are operating with the written records of a language which has ceased to be spoken, our chief witnesses are these records themselves. If the records happen to be inscriptions on stone or metal we are probably dealing with immediate copies of the original, and by a detailed study of all the forms found in the corpus as a whole, we can arrive at a satisfactory text which faithfully represents the original and avoid transcriptional errors. If, on the other hand, we operate with texts which are to be found in hand-written exemplars, our first task is to arrive at a satisfactory text by collation of available codices and application of textual criticism to constitute that text on the basis of all the codices which are our present witnesses to the original text. Here, then, is introduced a further consideration which goes somewhat beyond the immediate range of a linguist, but is a necessary part of a philologist. Unless out texts have been authenticated by a careful study of the manuscript tradition our lexicon will prove to be of little value.

Moreover, since we are operating with a writing system which has been traditionally associated with the particular language, we have to devise ways and means of judging their phonemic and phonetic values in order to know the true values of their central and peripheral sub-systems constituting the language. This involves a laborious study and analysis which is more difficult than that relating to a living witness who is able to react immediately and correct us if we go wrong. On the other hand there is a saving advantage that the codices are consistent at every point of enquiry, since however often we refer to a given instance at a given point there is no possibility of variation which often goes with a living witness who can react to your questions. This is a branch of study which must form part of the competence of a good lexicographer, although it is not an integral part of linguists. To this extent, at any rate, the old school philologist is better fitted as a lexicographer than a modern linguist who has not acquired the discipline of textual criticism.

1.10. The three-fold classification of language as emanating from the concepts of IDIOLECT, DIALECT and STANDARD LANGUAGE have given us specimens of lexicons at each level. The linguistic description of some of the unrecorded languages in India or the United States (chiefly the American Indian languages) involve a lexicon which functions primarily at the idiolect level. In the case of dialects, several idiolects have been drawn for the purpose of a lexicon. The standard

language of course is illustrated primarily in a literate society where it is the form of the prestige language as affected by the educated classes and in polished literature. Parallel to this is the system of building up lexicons for the usages of (a) a single author, (b) or authors belonging to a particular period of time or to a particular region or area. Finally the history of each lexeme as illustrated during the whole period for which the records of a language exist gives us a lexicon of the language during its whole life-time.

Idiolect and Dialect Lexicons often illustrate social and regional dialects. The ranking of society and the amplitude of differences of such social and regional dialects vary with each country and language. But if lexicons are based on this differentiation we are in possession of invaluable material which will help us not only in pure linguistic studies but also in the field of applied linguistics.

1.11. Since language manifests itself in space and time and functions as the most vital medium of communication among members of a given language continuum, it is characterised by these three coordinates. Since the continuum is defined by the language which it shares as its sole means of communication and cooperation, we can operate along the two axes of time and space, the continuum of speakers being always in the background, giving life and strength to the language. It follows, therefore, that the lexemes of a language can also

be defined in their space-time evolution. When we have a descriptive lexicon of a language we have a synchronous or diachronous treatment; in the first only those forms which are currently in use will appear as lexemes; in the second we have a repertory of forms which are current as well as obsolete, functioning along the time – and space–axes. Moreover, it may sometimes happen that while the lexeme itself is current, some functional significances which it carried at a given point of time have ceased to be current, and in these particular situations are replaced by others. It follows, therefore, that the space–time situation affects not only the free forms of a language but also the significances that these forms bear at the appropriate moment along these axes.

1.12. While we shall deal, in a different situation in a subsequent lecture, with the distinctive features in lexicography, it may be mentioned here that the principal criteria will be defined as (1) range, (2) perspective and (3) presentation. One aspect of range is the coverage in breadth and the coverage in depth. If L_1, L_2, L_3,......,L_n represent the number of lexemes in a lexicon, we may define n as the range (=coverage in breadth) of the lexicon so for as lexemes are concerned. If F_1, F_2, F_3,...F_m represent the functional meanings of L_1 in different situations in which L_1 occurs in the utterance event of the Language, *m* will represent another aspect of the range of L's which may be defined as coverage in depth. An ideally complete lexicon of the

language will try to extend the *n*'s and *m*'s to the largest dimensions possible in relation to its full range or corpus of texts. In the case of living languages the corpus is infinite or open; in the case of dead languages it may virtually be a finite or bound corpus. It follows, therefore, that within a closed or bound or finitely selected corpus of texts, a lexicon has a choice of limiting its *n*'s or *m*'s to any range from zero to the maximum possible within that corpus. Consequently each successive effort of editing lexicons will try to extend the coverage both in breadth and depth, while correcting previous efforts and deleting obvious mistakes.

1.13. Certain problems arise in extracting the lexemes of a language and arranging them in the form of a lexicon, a few of which only can be dealt with in these lectures. Naturally they will have to be the most important problems. They concern: (1) the identification of the lexemes and their arrangement in the lexicon (2) the identification of the functional meanings associated with each lexeme and their arrangement; if we are operating on the diachronic scale, the ordering of these significances not only in their space-time axes, but in relation to themselves, forms an important consideration. It has been demonstrated with certain European languages for which the material is rich and more adequate that the earliest meaning associated with a lexeme in its time axis is not necessarily the earliest meaning that that lexeme had originally. As Sommerfelt and Knudsen remark in their report to the Eighth

International Congress of Linguistics at Oslo: "If one examines most of these (the so-called historical) dictionaries one notes that the historical principle is not systematically applied in the exposition of the succession of meanings. Often the order seems to be a so-called logical one. It seems obvious that an historical dictionary must apply the historical view point also to the exposition of the different meaning of the polysemantic word. The oldest meaning is not necessarily the primary or the dominant one." The word for father, for instance, is a case in point. The original meaning in Indo-European was that of pater familias and not of physical fatherhood for which Latin has *geniter* corresponding to Sanskrit *janitā*. But modern European dictionaries for languages derived from Teutonic or Romance stocks denote that as the secondary meaning, preceded by physical fatherhood as the dominant or primary meaning. In many cases there is lack of the necessary data for a factual determination of the succession of the meanings. To a certain extent the principles of historical semantics many be of help. Formal criteria may also be an indicator. Thus in Modern Norwegian and Danish there are many derivations in *-ing* (cf English *-ing*, German *-ung*) which designate an action and also persons in a certain function. Here the formal criterion shows that the first meaning is also the first one in time. As a working hypothesis one may say that a meaning which contains an objective statement is older than one which implies a subjective valuation.

Such rules must be operated with caution, for it is impossible to establish mechanical rules of meaning development. If we are operating purely on the synchronic scale we would have to apply the structural point of view, distinguishing between the primary and secondary meanings from the purely synchronic point of view. Finally (3) the question of definition arises, whether the lexicon is unilingual or bilingual. Quite often synonyms are given in the unilingual system: e. g. the Dhātupāṭha of Pāṇini as presented by Bhīmasena gives the definition of the very first verbal base of the First Conjugation as *bhū sattāyām* while in the Second Conjugation the base *as-* is defined as *asa bhuvi*. Such examples can be multiplied. In such situations we may clearly remember the direction of Patañjali under the Vartika "Vyākhyānato viśeṣapratipattiḥ na hi sandehād alakṣaṇam", and occasionally a counter-example or negative definition may serve the purpose better.

———

LEXICOGRAPHY

II

2.1. The lexicon is an inventory of the free forms of a language arranged systematically, and against each form are shown their functional load of meaning in each distinct meaningful situation. If we confine our lexical activities on the synchronic level these meanings or the meaningful content of each lexeme will have to be arranged in a purely structured manner distinguishing between primary or dominant and secondary or derivative significances. Related to the dominance is also the question of frequency of occurrence of the particular situation in the whole range of the language. We may assume for practical purposes that the most frequent occurrence of a lexeme in a particular meaningful situation defines its dominant meaning content. In general, the history of a lexeme in a historically attested language shows that the general trend has been towards extension of the range of its meaningful occurrences along the time axis while its dominant or secondary meaning contents may change along that axis. Just as lexemes may be replaced by others in a given situation or context, either through natural evolution or through nonlinguistic factors coming into play, so also the meaning content of a lexeme may be enlarged or restricted in the course of unremitting transmission through successive generations.

The relation between a lexeme and its meaning content is arbitrary and is governed by the convention which obtains in the particular speech community. But once the lexeme comes into being with the particular conventional meaning content associated with it, it becomes subject to the natural changes which affect language in its different sub-systems during the course of continuous oral transmission from generation to generation. The manner in which the lexeme changes along with its meaning content can be studied if we have sufficient data to work upon. This is the principal function of historical semantics. On the other hand the same convention which is arbitrary but which in a given society associates a particular sign with a particular meaning content may on occasion function, under exceptional circumstances, to associate it with an unrelated meaning content, and start a new tradition. Some of these factors may be purely non-linguistic. Thus, in the absence of any indication of the exact meaning in a given context, as for instance, the want of an informant who speaks the language, and who can perhaps explain it, recourse must be had to various factors which must depend naturally on the situation concerned. We may have commentaries on the original text sufficiently near to the time of the text commented upon ; in this case we may accept the definition given by the commentary; there might be early glosses which might be taken note of, both as meaning definitions and as original lexemes. Occasionally we have the parallel

development in a cognate language which may to some extent indicate the process of development. Let us consider, for example, the Sanskrit words *aṃhriḥ* = *aṅghriḥ* foot, root of a tree and *pādaḥ* which has the following meanings: "foot; column, pillar; wheel; foot as a measure = 12 *aṅgulas*; root of a tree; foot of a hill or mountain; a hill at the foot of a mountain; the bottom (*dṛteḥ pādāt* = from the bottom of the bag); ray or beam of light (considered as a foot of a heavenly body); a quarter; quadrant of a circle; a verse or line; caesura of a verse." While both these are synonymous in so far as (a) foot and (b) root of a tree are concerned, the second has developed a number of new meaning contents which are absent in the first group. We may conclude from this that synonyms do not necessarily cover all the meaning contents of each other; they can be substituted for each other only under certain limited situations, and in so for as those situations are concerned they are identical, but completely different in the remaining situations. In a living language a speaker does not have to study the synonyms and substitute them in such situations when a certain style is being affected. We have examples in Sanskrit and other languages of writers whose feeling for the language has developed only through study of lexical material committing such unpardonable substitutions. Some of these have perhaps succeeded while the majority will go into limbo.

Now if we compare Sanskrit *pādaḥ* with English *foot* which has developed the following

meanings: "termination of leg, beginning at ankle; infantry (foot and horses); end (of bed, grave etc.); part of stocking etc. covering feet; metrical unit, with one syllable accented; linear measure of 12 inches; lower part, base, bottom", we notice that a larger group of meaning content is common to both. Comparative linguistics indicates that the Sanskrit *pāda–* and English *foot* are different realisations of an Indo-European lexeme. The main difference is that in a secondary sense Sanskrit *pāda* indicates a section or chapter of a book, which is lacking in English. On the other hand in the meaning content of infantry Sanskrit uses a derived form from the same verbal base as in *patti–*, *pattin–* and *padāti–*. It is not necessary that this situation should be reproduced in every single instance where cognate languages share inherited lexemes. But there is enough evidence to show that if the lexemes exhibit a one to one correspondence in their grammatical and phonological systems the chances of their sharing at least one inherited meaning content are high. In this sense cognate material from related languages can often help restore a possible meaningful content if the parallelism is fairly close enough. Recent research, for instance, in Old and Middle Iranian languages based on the Turfan and other material have thrown considerable light on the meanings of certain rare Vedic words which are almost ***hapax legomena.***

Since synonyms are only partial equivalents, a lexicon, if it is to serve the purpose of a complete reference tool, must not only give the definition

with or without the use of the synonym, but must also indicate the particular situation in which the lexeme bears that meaning-content. This constitutes the context in which the particular significance is illustrated. This is particularly required in polysemantic words which, in the course of historical transmission, have acquired a fairly heavy load of meaning content.

Quite often, when one operates with a single language, it happens that a given lexeme exhibits a variety of meanings which cannot be structured at the synchronic level. Let us take the example of modern Nepali *kāṭhi* which has the following two unrelated meanings: (1) saddle, pommel and (2) a sponger (fem.). The question here is whether we should consider these two meanings as belonging to a single lexeme or to two separate but homophonous lexemes, Without the aid of comparative etymology it is difficult to settle this issue; in the case of Nepali we know that in the first sense the Nepali lexeme corresponds to Sanskrit *kāṣṭhikā* with cognates in Hindi *kāṭhī*, Panjabi & Lahndi *kaṭṭhī*; in the second sense it corresponds to Sanskrit *kaṣṭaḥ*. We are therefore enabled to separate the two homophones by having two separate entries in the lexicon, distinguished by 1 and 2 prefixed or suffixed, either as running numerals or as supra- or infra-linear numerals according to the style we prefer. In the case of Nepali *kāg* (1) crow, raven and (2) cork, stopper, or in the case of *gol* (1) circle, crowd and (2) charcoal, the lexemes are loans from English

cork and *coal* with reference to the second meanings indicated but are inherited from Sanskrit *kākaḥ* and *golaḥ* respectively in the first meaning.

In the case of Nepali *ḍālo* (1) basket and (2) branch, bough, there would by some difficulty in identifying the lexeme. The question is whether the two meanings are structured as of a single origin from the same lexeme or have two homophonous lexemes fallen together. The meaning could be structured if the reference to branch or bough is to the bamboo reed. But apart from this we can look to collateral evidence and we discover that in the first meaning content Marathi has *ḍal* and Gujarati *ḍale* and for the second meaning content Gujarati corresponds with *ḍaḷ*. Again the comparative phonology of New Indo-Aryan languages indicates that Marathi and Gujarati intervocal single dental *-l-* corresponds to a double *-ll-* in the Middle Indo-Aryan stage while a retroflex intervocal *-ḷ-* corresponds to a single dental *-l-* in Middle Indo-Aryan. On this evidence we link up the first meaning with Middle Indo-Aryan (Prakrit) *ḍalla-* and the second with Prakrit *ḍala-* and so distinguish the homophones in the lexicon. In most of these cases we have taken recourse to comparative etymology.

It may some time happen that within each language the derivational forms from a single morpheme which is itself the result of convergence of earlier differentiated bases may be homophonous;

but they are to be indicated as separate entries in the lexicon.

The general practice may be summed up in the two following statements:

KITTEL: Kannada homonyms have been separated under two, three or more heads. This has been done not only when they clearly belonged to different roots, but also when there existed any doubt whatsoever as to their relationship: e.g, [2]*aṭṭe* 'a length' but [3]*aṭṭe* 'a handless trunk, especially retaining vitality'; [9]*aḍe* 'a sticky or clotty mass, lump as of dates or tamarind' etc.; [10]*aḍe* 'cowdung'; [11]*aḍe* 'mud' etc.

Tamil Lexicon: Homonyms of diverse origins are treated as distinct words.

2.2. Lexemes are sometimes identified by wrong division of utterance events. A classic example of this may be seen in Sanskrit *vidhavā* 'widow' which is a derivational form from the verbal base *vidh–*, as may be seen from the Indo-European cognates found in Greek *heitheos*, Latin *vidua*, Gothic *widuwo*, Germanic *wituwa*, *witewe*, *witwe*, Anglo-Saxon *wuduwe*, *widewe*, modern English *widow*. By wrong division we have *vi-dhavā* 'destitute of a husband' whence *dhavaḥ* 'a husband' as noted in Bhāgavata Purāṇa 1, 16, 20 or 'a man' Naighaṇṭuka ii,3. Indeed, not satisfied with this new lexeme occurring by itself as a result of wrong division, Sanskrit commentators are not wanting who see it in forms like *Mādhava*

(originally a secondary derivation from *Madhu*, but here rendered as *Mā-dhava* 'lord of Mā or Lakṣmī) and *Umā-dhava* 'Lord of Umā'. Later lexicographers and commentators in Sanskrit have been responsible in identifying a fairly large number of such lexemes resulting from artificial or wrong division of complex morphemes.

2.3. There are instance of lexemes which, like morphemes, are bound, and appear only in combination with other morphemes, usually as non-initial. In the usual lexicons there is no entry for such lexemes since they never appear in the free form. But since an ideal lexicon is a full repertory of all the morphemes or meaningful forms of the language even bound morphemes have to be included. Apart from the derivational and other suffixes or affixes or infixes, the lexicon should include bound lexemes also. An instance of this may be given here. In Sanskrit we have derivational forms like *haiyaṅ-gavīnam*, *āśitaṁ-gavīna-* etc. from which we can subsume a lexeme **gavīna-* which is bound so far as Sanskrit is concerned (in opposition to forms like *gāvī*, *goṇī*, *gopotalikā* etc., derived from the base *go-* and considered by Patañjali as Apabhraṁśa forms, i.e., forms not recognized as standard acceptable to the śiṣṭas). This lexeme has been inherited as a free lexeme in one of the Indo-Aryan languages of the modern period as I have shown elsewhere. It would, therefore, be necessary to have an entry in our lexicon under *gavīna-*, for instance, men-

tioning that it appears as a non-initial member in compound lexemes or morphemes.

2.4. This raises another important consideration in regard to the place of lexemes in the utterance event when such free lexemes may appear initially or non-initially. It would certainly be of much use to linguists and others if a comprehensive listing of compound lexemes is available of which a given free lexeme is a non-initial member. The best way to do this is under the free lexeme, and indicate there merely a list of lexemes with which it forms such composition as a non-initial member. To some extent this has been done in the Petersburg Lexicon of Boehtlingk and Roth. The only major lexicon in the Indian field where this is being done systematically is the Critical Pali Dictionary of Trenckner.

2.5. When we are dealing with a language which is still living, but which has recorded history of a fairly long period, with a given writing system, one of the aspects which is expected from the lexicographer is the actual phonetic value of each lexeme in the approved standard received by the educated classes. In the case of purely dialectal forms the local pronunciation is to be indicated. Graphemics lags behind phonetic or phonemic change realised in the language, and to the non-native speaker the traditional spelling conveys a wrong idea. It is, therefore, necessary that where such lag exists the lexicographer should

once again function as a linguist in the field eliciting information about the exact values of the lexemes in terms of articulatory phonetics or phonemics if the phonemic analysis of the language has been carried out and statements prepared on the repertory of such phonemes. In our Indian conditions this is particularly required so far as all Modern Indo-Aryan languages are concerned and perhaps in the Dravidian group Tamil certainly and others perhaps to a less extent, require this aid correctly to identify the lexeme in current speech.

2.6. Since etymology cannot be completely ruled out in a lexicon where, as we have seen above, its help is necessary in separating homophonous lexemes converging together through linguistic change affecting different earlier lexemes, the question of its inclusion often becomes vital to the lexicon maker. Very often this takes the form of noting down the cognate forms in related languages. While this is an essential feature, a true etymological study would go beyond mere listing of cognates and establish by comparative reconstruction the earlier or earliest common form of which they are different realizations. In the case of Modern Indo-Aryan languages this would lead up to some available form in Old Indo-Aryan or in a certain number of limited cases to reconstructed forms. TURNER's Nepali Dictionary, for instance, shows that in a corpus of approximately 27,000 lexemes listed there, nearly 5000 show the existence of cognates in other NIA

languages, and comparative etymology shows that approximately 20% of these lead to IA reconstructions. Perhaps a more thorough search of OIA lexemes hitherto escaping the attention of earlier lexicographers may discover a number of actual current forms to replace these reconstructions. In the case of Dravidian languages we will probably have to go back to Proto-Dravidian, although Tamil exhibits forms nearest to this stage. It is not necessary to go beyond, in general, the immediate ancestor, as OIA for NIA; very rarely is it necessary to consider Indo-Iranian or Indo-European forms except where OIA tradition is not available for verification due to incomplete recording of material. Attempts such as those which actually exist, linking up Pali with Indo-European are good pieces of exercise, but throw little light on linguistic evolution, since we shall be comparing strands at different stages of transmission and missing the intermediate links which exist and through which they have actually passed.

A specialized type of lexicon has come into existence for giving the etymology of either a single language or the comparative etymology of the whole family. In either case we have to list the cognates in the same line of descent and link them up with earlier forms current in the previous lines of transmission, step by step, leading ultimately towards the reconstructions of the most ancient archeform of which they are subsequent realisations in their individual lines of transmission. The procedure in the first case is

determined by the repertory of lexemes in the individual language selected; that in the second case is governed by the totality of lexemes in all the member languages, and in consequence will cover a much larger ground in depth as well as in breadth.

2.7. One consequence of the etymological approach is the discovery of lexemes which have been taken over as loans from (a) unrelated languages, from (b) related languages at an earlier or parallel line of transmission or (c) inherited from early loanwords at a previous level of transmission. The problem of loanwords is fraught with great difficulty, particularly when one deals with the Indo-Aryan, Dravidian and Austro-Asiatic families. There is a very influential school of thought which sees in the development of Indo-Aryan phonology within the Indian sub-continent the gradual influence of the speakers or speech habits of Dravidian and Austro-Asiatic, particularly in the development of the retroflex series of stops and laterals. A study of related phenomena has led to the positing of an Austro-Asiatic sub-stratum in both Pre-Aryan and Pre-Dravidian in India. Thus KUIPER in his Proto-Munda words in Sanskrit attempts to trace a number of Vedic expressions to Proto-Munda. Earlier Sylvain LEVI and PRZYLUSKI attempted to show the penetrating influence of Austro-Asiatic on Sanskrit, and BLOCH on Proto-Dravidian. This is not the place to judge the value of these contributions or the methodologies employed in

arriving at such results. I have shown elsewhere that in unravelling these tangled skeins in the development of the linguistics situation in India we must depend very largely on the first principles of the three-dimensional character of linguistic growth; if these are not used to check every step of our reconstruction we are likely to fall into grievous error.

The prehistory of such words, particularly suspect of being loans, can be studied with advantage if we can link up a group of them connected with some art, craft and technique. In such a situation we have, beside the regular linguistic evolution, collateral evidence from social and anthropological history, which will correct each other as we proceed along.

2.8. When a lexeme is associated with particular meaning content, and these significances have been registered, the lexicographer is ready to arrange them. The historical lexicographer must arrange them in the chronological order of their development, as far as he can ascertain it. When, as in most cases, there are not more than a few, say four or five meanings, usually there is little difficulty in arranging them. But when there are multiple meanings it may require considerable research to discern the true order of development. It has already been pointed out that it is not sufficient to place the senses in a simple series according to date of appearance. We must acquire the fundamental aspect of historical semantics and

the principles which govern the proliferation of meanings. It may, however, be mentioned in passing that meanings spread out from a central significance which may or may not be preserved in the particular lexeme but may be inferred from an earlier language or may be entirely lost. Each of these derived meanings may become in turn a centre from which further secondary meanings may develop. Meanings may consequently be related to the central core directly at each step, or by successive links leading back to the core, each of the derived link points serving as focal points of further development.

Different methods have been used in indicating the ordering of the meanings. Either Roman or Arabic numbers singly or in combination, with further use of alphabetical sigla, have been used to construct the edifice of ordered significances. In the case of popular words bearing an unusually large load of meaning content, the material may spread over hundreds of columns and figures. Examples of this may easily be seen in the New English Dictionary published by the Oxford University Press.

2.9. It is only a lexicographer who becomes conscious of the polysemantic lexeme or word when he begins to organise his basic material. In ordinary usage we are hardly conscious of the fact that the words we use have such a wonderful meaning content. We are governed by the social needs of language communication and as each occasion arises

the necessary message is automatically composed and uttered without conscious use of the stock of morphemes which the individual stores in his memory-complex. It is perhaps in formal composition and an individual style of speech that words acquire the specialized shades of meaning which are associated with them in literature.

In no language can we discover the happy situation of one lexeme having only one significance. In all extant languages whose existence can be verified or inferred from written records, many words have several, and at times a large number, of meanings. The reason for this development of several meanings of one word may be connected with the advance of civilisation or culture which creates a need for means of expressing many new ideas, naming new products, expressing new kinds of action, etc. Two ways are employed for this purpose: (a) a few of the important lexemes may be borrowed as loanwords from languages which have words for these new experiences; (b) the more usual method for increasing the range of lexeme-meanings is through addition of new meanings to words already in the language. Thus the existing lexemes begin to serve a double purpose and gradually acquire a larger field of usage. An unravelling of this process in its time and space axis will illuminate many dark corners of history for which there is no other record available.

2.10. If the lexicographer is operating with a unilingual lexicon he has the additional responsibility of supplying accurate definitions ; for as pointed out earlier, partial synonyms, even though they may be more frequent than the particular lexeme whose content they bear in the particular situation, hardly convey adequately the full range of that content. It is all the more necessary in this case to support the meaning by a suitable but accurate definition. Such definitions are possible for a great part of the vocabulary because the content of the word is complex and may be resolved into different elements. The definition should really cover the content of the word and not contain either more or less than the content; such a definition must be able to take the place of the word in the chain of utterance without any change in meaning. Whether it is the intention of the lexicographer to give the citations or not, he has to collect them in order to arrive at the meaning content of the particular word or lexeme, and this will help him in coining a suitable definition which may replace the lexeme in that particular chain without changing the meaning of the utterance as a whole.

When we operate at the bilingual level and the meanings are expressed in the second language, we adopt a process of translation both of the synonym and of the definition in such a way that the original pitfalls indicated in the case of unilingual lexicons may also be avoided in the bilingual lexicons. An accurate translation gives a

dual command of two languages and makes for refinement of speech in both directions.

2.11. It may not be out of place here to indicate a few practical suggestions for the compilation of lexicons. That for a language which has no records and no writing system involves a complete linguistic analysis and building up of a corpus of utterance events elicited from the native speakers of the language. The lexicon here is one of the end results of field work in descriptive linguistics. To some extent the same methodology has to be adopted in relation to the dialectal lexicons; selected dialects will have to be analysed and given a proper writing system (whether phonetic or phonemic) and the corpus built up gradually with the aid of the native speakers of these dialects.

In the case of standard literary languages whose lexicons we wish to compile by applying historical methodology, we have first to determine the size of the corpus which should give us our material. Along with the size we have also to set the two limiting points on the time axis. The coordinates along the time axis will give us our historical perspective ; we must now determine our range which, as pointed out earlier, shows coverage in breadth and coverage in depth. The number of texts falling within the time axis fixed by predetermined coordinates will generally govern this breadth; but their selection on the basis of (a) regional distribution and (b) type of composition and (c) technical character will contribute

to some extent to the coverage in depth also, as this choice will determine the inclusion or exclusion of some types of lexemes and their meaning contents. Too many texts belonging to one period of time will certainly determine the period character of the lexicon; similarly too many from a particular region will lay heavy emphasis on regional vocabularies. The choice must be exercised from the particular perspective in mind.

Once the corpus of texts has been finalized, the best way is for competent readers to go through each and underscore the words which should be considered for inclusion. We have two processes open for us here: (1) it may be our intention to be absolutely inclusive and omit nothing; in this case we are including everything irrespective of the number of repetitions of an individual lexeme in a given meaningful situation; this may be called the Thesaurus method and every item will be recorded; (2) we may be selective in our method, while trying to be as comprehensive as possible; in this case we shall, as far as humanly possible, eliminate repetetive entries in similar situations while recording one occurrence at least of each word in each of its distinct meaningful situations. The marking of the text has a special meaning only in this case, since we are not attempting an index verborum. It would be a good practice if the text marked by one scholar is checked by another to ensure that the accepted principle of selection has been more or less rigorously followed by the first scholar.

After the marking of texts has been completed we may now pass on to the next stage of extracting this material for our proposed lexicon. Obviously we must utilize suitable intex cards made of strong durable material, preferably with rag content designed to withstand long and constant handling. The basic principle is that we must use one card for one lexeme in one meaningful situation. The card should provide at the top sufficient space to write the lexeme in its basic form, with the particular meaning content written against it. The specific situation in which the lexeme occurs in the speech stream may then be noted in the central part of the card, followed immediately underneath in the space reserved for it the exact reference to this citation, with specific indications to enable a reader to find it in the text any time he wishes to consult the original. In a specific part of the card prominently there should be room to indicate the date of the text from which the material is being extracted. Since most of this work involves two kinds of competence:— that of a scribe so far as vocable, citation and reference are involved, and consequenly lends itself open to the kind of transcriptional errors we may expect in manuscript transmission; that of a scholar who has to determine the exact meaning content in this context and also the actual length of the citation which makes that meaning clear to a discerning scholar, each card should be completed by one scholar and subsequently checked by another before it can be accepted for the lexicon. Normally,

therefore, such a card should provide for the record of the original compiler and the corrector or checker.

When each work has been completed and checked the material extracted from it will have to be alphabetized and put into the proper rubrics in the Dictionary Scriptorium. Under each major rubric there will be a single lexeme, and under each lexeme there will be as many sub-rubrics as its meaning content requires. The particular card will go under this meaning sub-rubric in the major lexeme rubric and find its place in the appropriate time scale.

The work of editing which commenced from the reading of a text and arrived at the point where the material is alphabetized and rubricised, continues unabated. For in each sub-rubric will be found situations in which a particular meaning content is associated with the lexeme; in the original situation each entry was considered solely on the basis of a single text at a particular point of time; in the present situation we have a number of contexts in which the particular lexeme is associated with a particular meaning, and possibly by a collation of similar situations we may be able to refine the definition of meaning content or split them into further subdivisions. Examples of such study are galore in Vedic interpretation; possibly a similar situation may hold good for ancient Tamil literature. Philological and linguistic journals carry a number of research papers on the meanings

of difficult or unusual words, where this method of comparison of similar situations has resulted in precision of meaning definitions.

For the purpose of the lexicon we may distinguish between primary, secondary and tertiary sources. The original texts selected (which must of course be critical editions based on the collation of representative manuscripts) constitute our primary sources. If there are ancient commentaries which themselves are not included in the list of texts, these become our secondary sources; if not, they combine in themselves both primary and secondary sources. Translations of the text by competent scholars provide a tertiary source along with special studies of the work, particularly containing discussions about specific words. All these must be collated if we aim at a comprehensive sweep. No contribution, however small, can be neglected when we work on such a wide canvas.

What has been briefly indicated here is very essential for a lexicon which is being complied by corporate activity. A system has to be introduced if the different contributors have to conform to a given pattern and standard. Many of the lexicons which have to be attempted in India during the next plan period will have to be corporate efforts if justice is to be done to our literary heritage.

Universities will, of course, be in the forefront of this activity if we take note of the Report of the Blue Print Committee which submitted its findings to the University Grants Commission last year, with regard to the development of linguistic studies in our Universities.

LEXICOGRAPHY

III

3.1. Lexicon making and its problems are of interest to the scholar and to the educator, to the litterateur and to the practical man, to the student of an individual author of genius, and to the investigators of society and culture, and to the historian and to the contemporary student of world affairs. Language is indeed all-pervasive, and consequently language usage, and the lexicon which systematizes and records it, is of interest to all thoughtful men and women. The systematizing and recording of language usage is of course of particular moment to linguists who are professionally engaged in describing language and in tracing its history, and to philologists who devote themselves to the elucidation of the cultural content of a people's linguistic records. Language is a notoriously flexible instrument which registers culture change perhaps more sensitively than does any other phase of culture. Such changes may affect phonetics, vocabulary or grammar. It is primarily in vocabulary that they can be traced more readily, whether they be conceived as due either to developments within the culture or the effect of external stimuli. Many of these become apparent only when a lexicon has been properly compiled, and the perspective on which it is built up. We may, therefore, at this stage consider the

distinctive features in lexicography by enumerating the typological aspect of lexicons.

3.2. In an earlier lecture the criteria mentioned were: [1] range, [2] perspective and [3] presentation. We shall first deal with classification by Range = the volume and spread of the material assembled. We have already indicated that if $L_1, L_2, L_3, \ldots\ldots L_r$ represent *r* lexemes in the language, the larger *r* is, the wider the coverage. When *r* is sufficiently large we have a patently representative coverage, although we may not be absolutely certain of its absolute percentage to the ideally complete record of all the L's in the language. This coverage is in breadth, in the wide sweep we make for catching the lexemes or minimum meaningful forms and their combinations in forms within the corpus of the language. If $F_1, F_2, \ldots\ldots F_s$ represent the functional meanings of L_m the larger *s* is, the deeper is the coverage. We have thus two ways in which range is measured, [1] coverage in breadth and [2] coverage in depth. The first determines the number of independent lexemes while the second measures the wealth of meanings of each lexeme. Successive efforts at compiling lexicons will aim to improve on existing resources by addition to both kinds of coverage, and increase the range of the lexicon. Of course this need not necessarily be so except in a purely scholarly lexicon, as school and college and commercial dictionaries have only a limited aim.

3.3. To ascertain an ideally complete record of all lexical formations within a given language, it is necessary that we must have the complete corpus of the linguistic record in that language and from this corpus we must carefully extract all relevant forms by making *r* the maximum possible within that corpus. As mentioned earlier this is the thesaurus method and it is not generally applicable to a living language whose lexicon is unbounded. Normally most dictionaries or lexicons represent variously scaled abridgement of such a complete record of all lexical forms. If the work has not been done with care by a competent philologist the lexicon may contain obvious mislections or misprints respectively from manuscripts or printed works; spurious readings; and in the case of lexicons of modern Indian languages a fairly wide-spread category of new forms manufactured from the Sanskrit lexicon to suit nonce-formations.

3.4. We may operate with social dialects or individual texts or regional dialects; we cannot escape from this first criterion of range which affects them in the same manner as lexicons of the standard literature in the language.

The range may sometimes be determined by the frequency of occurrence of the lexical items in the stream of language or utterances. This especially occurs if the aim is to serve a particular class of society.

The range of a vocabulary may be delimited by the distance between points selected along the time axis, while its coverage in breadth may be increased, maintained or reduced according to the particular objects in view. One of the byproducts of this will result in the optimal representation of the entire vocabulary of a language historically attested by means of a series of chronologically bounded dictionaries.

Lexical analysis of dialects may be defined as involving a dimension of space, while the analysis of literary texts primarily entails a movement along the time axis. But the two are at no time mutually exclusive, since dialect-texts may also be available, and we may have to combine both approaches in a single effort. This is indeed the case with modern Indo-Aryan languages which, in their earliest forms exhibit clear dialect cleavages before the language emerged in some royal court as an accepted standard by the elite of that period.

35. We may classify lexicons by the number of languages involved. Initially in India we say the beginning of mono-lingual lexicons was followed later by the bilingual lexicon, the chief language of definition being the classical language Sanskrit. We may have for the modern lexicons in the Indian as in other fields, either mono-, bi-, tri- or quadri-lingual or polyglot dictionaries. The language of definition is one criterion for classification by distinctive features. The usual bilingual

dictionary such as English-French or French-English may be replaced by Hindi-Hindi-English which incorporates characteristics of the mono - and bilingual dictionaries; we may well conceive of new trilingual dictionaries during the third plan period of such type as Tamil-Hindi-English and various combinations of the same. In particular we may also conceive of technological multilingual dictionaries, since each regional language in India is busy compiling technological vocabularies of its own.

3.6. Another aspect from which classification by range may be examined is the degree of concen tration on lexical data, with particular reference to mixing of lexical information with encyclopedic details. This is particularly in evidence in the inclusion under proper names of historical or biographical vignettes, etc. Many lexicons, however, prefer to relegate such historical, biographical or geographical information, to appendices.

3.7 Under PERSPECTIVE we may take note of *the following:* [a] *with regard to the time axis*, the outlook may be historical (dynamic) or synchronic (static); [b] with regard to the basic arrangement, the sequence may be conventional (=alphabetic), semantic (=ordering by parts of speech or provinces of life), or entirely arbitrary (=chaotic); [c] with regard to purpose, the prevalent tone may be objective, preceptive or prohibitive or jocular.

It is hard to find among the many existing dictionaries of Indian languages a purely synchronic or static lexicon; what purports to be a modern lexicon mixes up the synchronic with much diachronic material, and given latitude includes a large section of the Sanskrit lexicon copied out without reference to the actual needs of the language. Similarly no dictionary on historical principles is available, although we have short lexicons for single authors or for particular periods of history in the development of the language. Both, to my mind, are now required, particularly with reference to the great regional languages. Nevertheless this distinction between synchronic and diachronic approach is a very significant factor in the typological study of lexicons. The arrangement of meanings will be different in these two cases, with reference to the same language: the synchronic lexicon will organize its meaning suitations in a structured manner without reference to the historical evolution of these meanings. The diachronic approach will have to arrange the meaning situations with relevance to [a] the historical appearance of each meaning situation and [b] the overall development of such meaningful situations on the basis of the etymological evidence, *i. e.* the prehistory of each form and meaningful situation.

3.8 Arising from the diachronic approach or perspective is the so-called etymological dictionary. This, as I mentioned earlier, may proceed along two complementary lines: [a] the etymologies may

be arranged under the headwords of lexemes of a given language and so be limited by the number of such lexemes; [b] they may be arranged under the total number of lexemes available from all the cognate languages and recorded under the oldest or archeform available or reconstructed from such correspondences of member languages. Arising from this type of lexicon we may also take note of studies of specific lexical strains: e. g. Portuguese loan words in modern Indian languages or Arabic or Persian loan words.

3·9. There are three contrasting basic arrangements of lexical items: alphabetical, semantic or casual.

The *alphabetical arrangement* is *certainly* the most overwhelmingly dominant and, although conventional, has acquired the appearance of naturalness to such an extent that an ordinary person normally associates this genre of the lexicon with such things as a telepohne directory, a mailing list or a classified catalogue. In its origin, however, the alphabetic arrangement came fairly late. In the case of compound letters such as *kṣ* or *jñ*, the traditional alphabets of most modern Indian languages place them at the end of the normal alphabet, and entries thereunder naturally came to be placed at the end, and not under the appropriate order of *k* or *j*, since graphemics provided special symbols for them.

There have been lexicons printed in the reverse alphabetical order, mostly as *indici rerum*

about which the late Professor Jules Bloch mentioned in his Furlong Lectures as a necessary tool for linguists, illustrating how such an aid assisted Latinists in delineating the history of certain formations in that language.

Remotely akin to the kind of lexicons noted above are rhyming dictionaries of which one at least in Tamil is known.

3.10. The semantic arrangement is probably the oldest; all the early Sanskrit and Tamil lexicons available may be grouped under this classification, the lexemes being arranged in a semantic link under specific ideas which themselves are classified by objective world phenomena.

3.11. Classification by parts of speech is also a well known device. Mention was made earlier or instance of the Gaṇapāṭha which listed classes of nouns of verbs or adverbs etc. according to its form-class in the language; similarly the Dhātupāṭha listed the verbal forms of Sanskrit.

3.12. Indexes to critical texts, historical grammars and other linguistic or philological studies including good translations often constitute a kind of lexicon with casual arrangement.

3.13. There are three contrasting levels of one of a lexicon: [1] detached, [2] preceptive and [3] facetious. The first of these is the only one which insures a strictly erudite or scientific approach. This category includes the great majority

of uni-and bilingual reference dictionaries and glossaries which show a *historical slant*. The lexicographer here is concerned either with the structuring of the meanings in a synchronic description of his lexemes or a historical arrangement which duly takes note of the historical *development of the meaning content*.

The preceptive or normative or didactic tone of a lexicon presupposes a refusal on the lexicographer's part to accept the basic facts of language as amenable to observation and his belief in an ideal state of affairs bearing the stamp of orthodox and appearing to him worthy of inculcation. One special class of lexicons which fall under this category are the remedial dictionaries which overemphasise the retention of traits of the prestige language [in orthography, orthology or syntax] to which the average student is likely to prove difficult.

The third attitude is occasionally displayed by even the most sober lexicographers, and we have a number of definitions given by Dr. Samuel Johnson which fall in this category. But so far as lexicons in India are concerned there is no single representative of this third perspective.

3·14. The third criterion we had in mind was classification by presentation. Perspective helped us to account for the broad policies of selection, arrangement and tone; PRESENTATION is used as a term for getting at the narrower preferences such

as typestyle, use of special symbols and abbreviations, inclusion of incidental grammatical remarks, latitude of definition, quantity of verbal documentation and graphic illustration, etc. We may, however, concentrate on the following four important points: [a] definition, [b] verbal documentation, [c] graphic illustrations [inclusive of maps,] and [d] special additional features.

The range of specificness and fulness of a definition may range from a single gloss or synonym to profusion of technical detail. A lexicon is a tightly built work of reference and there is a certain limit to this wealth of definition beyond which it is not prudent to go; similarly at the other end there is a different kind of limit, between which a lexicographer has to function.

In judging documentation it is advisable to separate the literary language from living dialect speech. Where literary usage is at issue, specificness of documentation may oscillate between the entire passage or a somewhat shorter segment of that passage or a bare reference to the line, folio, page, chapter etc It is for the lexicographer to decide which of these three modes he is to follow in a particular case. If the dictionary functions at the bilingual level the lexicographer may provide a running or exact translation of the passage cited.

3.14. In pure dialect dictionaries the formality of explicit documentation may ordinarily be

dispensed with. But a conscientious dialectologist-lexicographer must avoid coining his own citations and depend almost entirely on the corpus of dialectal material he has collected.

In situations where scientific definition or an abstract description is required, very often the inclusion of drawings or sketches serves as a quickener of interest and makes definitions concrete. The photograph embellishes it with authenticity. The inclusion of any of these depends entirely upon financial considerations; their utility, however, is unquestioned.

3.16. Among special features we may include abbreviations, phonetic transcription, transliteration. Abbreviations, within a certain range, have particular use; but in a historical lexicon where citations have been collated from over 1000 texts they will prove to be of little use, and their inclusion may defeat the very ends for which the system is introduced. Phonetic transcription or systematized transliteration or phonemicisation may be another special feature designed to assist the non-native consultant of the lexicon.

3.17. We mentioned earlier that a lexicon is a repertory of lexemes to be found in a language in their different meaningful environments which the lexicographer codifies and presents in a systematic manner as a tool for the person who is required to operate with the language. It is primarily a reference work which supplies at a

moment's notice the specific need of a scholar who seeks information therein. These needs may have something to do with the exact graphemic representation of a word or its specific meaning in a given context or the finding of a suitable expression for the particular situation that he has in mind. There might be hundreds of other reasons for which a lexicon may be required as a reference tool.

There is one aspect, however, which deals with the cultural history of the language community in which the materials included in a scientifically executed lexicon may prove to be of immense use. While literary documents may give graphic pictures of the civilisation and culture of the period which they depict, they may omit reference to the most common-place things. If we are to reconstruct the daily life and business of the common people from sources other than actual descriptions a study of the 'biography' of words or lexemes may throw a considerable amount of light. Let us imagine for instance a Proto-Dravidian community. From a study of the lexical elements found in actual use or currency at any stage of the attested Dravidian languages we find that proto-Dravidian had words for milk and some milk products; of these all, other than those signifying "butter" and "clarified butter" appear to be pure Dravidian; the word for butter and clarified butter is an exception, and it shows relationship with OIA *sneha–*. Anthropology tells that there are pastoral communities which

live on milk products and who are supremely unaware of the use of butter or curds, and particularly butter. Some of these are to be found still in the domain of Indo-Aryan speakers although Indo-Aryan languages have this word and in general know the use of butter. On this analogy are we to conclude that so far as existing evidence goes proto-Dravidians did not know the use of butter, though they have specific Dravidian words for milk, curds and butter-milk? Here is problem which linguistics and anthropology pose on the basic study of lexemes to be found in Dravidian and Indo-Aryan. They indicate a field of enquiry which is worthy of further pursuit.

Again we find that while Sanskrit has *aśva* recorded from Vedic times in its entire history, modern Indo-Aryan languages have replaced it by the later *ghoṭaka* which is first attested in the OIA language by about the 2nd century B.C. It is this *ghoṭaka* which is reflected in the various Dravidian words for 'horse" and gives us a reconstructed proto-Dravidian from **ghutr*– with initial velar aspirate. *Ghoṭaka* is not Indo-European nor is *ghutra– Dravidian. We have to assume a common borowing by both speech-communities from some other source. There is no historical evidence that the Aryans did not know the horse and imported it from outside. Something must have happened between the 3rd and 2nd century B. C. which made it necesaary for India to import horses from a particular source which contributed *ghutra - as

ghoṭaka in OIA and various forms in the literary Dravidian tongues. This linguistic evidence may give a clue to some kind of culture contact which led to acceptance of a loanword in OIA when it had already inherited forms for the horse.

We might, for instance, find out from our lexicon built on historical principles what expressions exist in the language during its particular history, of the means of communication such as pathways and byways, roads and alleys and all kinds of lanes. On the basis of the lexemes recorded in these meaningful situations and the dates during which they were current, and the first recorded instance of each, we may be able to find out more successfully than through written history, the kinds of communications existing in villages and cities and towns, and within a village, between different residential areas. This will give us a more detailed picture than we can glean from recorded history. If we can link this up with the etymological history of lexemes existing in the members constituting a single family related to the language in question our reconstruction almost parallels that attempted by the field archaeologist on the basis of culture objects he unravels in proper stratigraphic layers of which accurate records are kept by him. We may extend this to parts of a residence and reconstitute it at any point of time on the basis of lexemes current during the particular period. The linguist utilizing a scientifically constructed dictionary on historical principles may select his material bearing on the cultural life of

the community and reconstruct fairly accurately the course of history on such aspects. Not much has been done in this direction so far, but a consistent study linked up with field archaeology and anthropology will lead to an all-round development. Each will serve as a mentor on the other discipline and all the three in consonance will be able to achieve more accurate results.

3.18. It is not possible within the space of three lectures to indicate even in a nut shell all the important leading features involved in the making of lexicons, whether at synchronic or at diachronic levels. I must bring these to a close by pointing out to some of the important aspects of lexicography which have an immediate importance in our cultural history and in the development of our country as a single nation.

In the first place we see that none of our state or regional languages have adequate lexical aids. I speak from personal experience from a scholar's point of view. The best lexicons produced during the last century marked the happy co-operation between Western scholars and Indian scholars versed in traditional lore. It was primarily during the present century that the full benefits of the Western Educational System could be seen in the coming into being of competent scholars who derived their inspiration from traditional lore but their competence from modern education. Even so, our regional languages still stand in need of adequate lexical aids. Those existing are a

hyotch-potch of *selections* from partly literary sources and partly from dialectal sources, but by and large inadequate in regard to both. Even such exceptional works *like the Sabdasagar in Hindi* or the Madras University Tamil Lexicon fail to satisfy if judged by these standards. The entire field of scientific lexicography [=lexicology] awaits the willing scholar.

How then should we proceed? On the one hand there is the very large question of winning our substandard and dialectal forms of regional languages which contain a unique wealth of lexical elements which have not been recorded at all. This will have to be done systematically by field techniques and everything that can be recorded should find a place in these investigations. This will involve all the processes of linguistics field work and entail the full record of all linguistic forms and utterances in each dialect in the whole of the region. On the other hand there are a number of unrecorded languages and dialects which need the linguists' immediate attention. The Government of India have been proposing to establish a new Linguistic Survey, and one of the most urgent tasks is to run a race between complete scientific analysis and recording and final extinction. All this cries for competent liguists who should become creative artists by working through this material and provide the requisite lexical aids. The third aspect is the lexicon of our great state languages on historical principles. Kerala University is well in the final stages of a dictionary of Malayalam on

historical principles. The Kannada Sahitya Parishad, with substantial assistance recently promised by the Mysore State Government, has undertaken a similar dictionary of Kannada; and now your University has accepted in principle the need for such a dictionary for Tamil and is seeking the required assistance from the State Government and the University Grants Commission. I am not fully aware of the kind of dialectal material that the Kerala University project has incorporated and I know less of the Kannada Sahitya Parishad scheme. I may therefore, draw your attention to the fact that side by side with extracting your material from textual sources, you must try and collect full dialectal material; in fact you should plan to bring out in advance of the major work comprehensive dialectal lexicons of spoken Tamil. Each of this could be done by a single scholar, under suitable supervision, which is amply available in your University. Perhaps when the linguistic survey is established one unit may be functioning under the auspices of your University, if the suggestions contained in the Blue Print Committee report are acted upon by the Union Government jointly with the University Grants Commission.

You may perhaps raise the question why such lexicographical activity should be undertaken at this juncture. It is a valid question which requires an unequivocal answer. During the British regime our national languages were replaced by Engish as a medium of education and language of administration at a time when new scientific

discoveries added considerably to the stock of lexemes of all European languages. Two hundred years of such development and the lack of support for our languages have left them in a comparatively poor state of development; today they are proving inadequate fully to meet the challenge of this atomic age. Our needs have outstripped our equipment in linguistic forms, and like the state of education and our achievement in various fields of science in general we are almost two centuries behind, with little in our favour to catch things up. Indeed the situation is such that even the most optimistic among us will have to yield that the chances of our catching up with more advanced nations are slim, while on the other hand the certainty is great that the gap may even widen. This is indeed a challenge to language scientists who have been called upon to assist the nation in this uphill task. Efforts have not been wanting to CREATE new forms artificially without regard to the fact that words cannot be created ad hoc on this scale and be accepted by the community particularly in a secular democracy. Patanjali has warned us that words cannot be created for our use on the analogy of pots and vessels which can be so manufactured. Dr. Raghu Vira's efforts were described by German newspapers as a Manufactory of Words in a purely appreciative spirit; he has now gone from this to the publication of the Satapitaka, and for the time being the manufactory has come to a stop. In this zeal for creating new words in each region and by pedestrian scholars the

major fact has been forgotten that our dialects and unregarded languages are rich in lexical forms which have eluded lexicographers, and even a plodding effort of short duration is capable of winning hundreds of such forms, not to say thousands. It is a curious fact that such dialectal forms often run across language areas, and may, *when suitably won and employed* in standard languages, bring divergent languages together. The linguistic wealth of India has never been estimated much less counted. Grierson, at the end of his long and profitable career as a linguist in charge of the great Linguistic Survey admitted that Indian languages could easily compare with the best European languages in the wealth of their construction and in their lexicons and were capable of giving expression to the most refined ideas in every walk of life including modern science. Their inherent capacity is not less than that of the most advanced languages of the East and West. But we must equip them with the necessary sinews by winning this linguistic wealth and reburnishing it by putting it into active currency in the standard regional languages.

3.19. There is the other related question of bringing together the different speakers of the regional languages for the purpose of developing the offcial language of India by endowing it, without interfering with its individual genius, the forms, expressions and style of these languages. How can this be done when we have only a small percentage of the linguistic wealth in the country

available to us? Dialectal forms which run across language areas and unite speakers with different affiliations are more certain of achieving a certain unifying effect than incorporation of language forms across such borders. Not quite distinct from this is the other question of how the regional languages themselves may get enriched with these elements in such a way that they can share many recognition vocabularies drawn from such dialectal forms which tend to bring together distinct language areas.

3.20. It is seven years since the conference of Linguists and Educationists met in Poona under the auspices of the Deccan College to discuss the fundamental needs of Indian Linguistics and to devise ways and means of imparting modern linguistic tools and methods to Indian scholars. It is six years since the series of summer schools and autumn seminars were started, and over 1000 scholars have received various degrees of training in different aspects of linguistics. Today we are in a much better position than seven years ago so far as staff and equipment are concerned. Several Universities have departments of linguistics, and linguistics has received proper recognition both from the Universities and the University Grants Commission. There is every possibility that various governmental and other agencies like the Sahita Akademi may embark on ambitions projects of bringing out trilingual and quadrilingual lexicons during the third plan period. This period may also see the initiation of the Linguistic Survey, and a

new Central Academy of Indian Languages may come into existence as recommended by the Official Language Commission. Looking to all this tremendous development we may safely assume that present-day linguists will be called upon to bear a fairly heavy responsibility during the ensuing years. In the language of ancient wisdom I may call upon you with the admonition: "Arise, Awake; Stop not till the Goal is reached!"

3.21. I am grateful to your enlightened Vice-Chancellor who is unfortunately not here, and to Professor Meenakshisundaran for extending to me this cordial invitation to come here and share my experiences and joys with you in delivering a course of three special lectures on lexicography, particularly in view of the fact that you will shortly be inaugurating that great project of a Dictionary of Tamil on historical principles. It has been my privilege to have inaugurated at the Deccan College the monumental project of a Dictionary of Sanskrit on historical principles which is basic to all linguistic research, and this was an opportunity to be associated with another such project which is rightly being undertaken by this great University. Fate seems to bring me in association, I may perhaps be permitted to say close association, with your University in matters pertaining to linguistics. I had the privilege of meeting the Chairman of the University Grants Commission here three years back and discussing the development of your Linguistics Department at the time he personally announced his approval of that

project and sanctioned grants; now I am meeting you as a full-fledged University Linguistics Department on the eve of your great undertaking which may perhaps determine the specific character of your department. For the past 6 years Professor Meenakshisundaran and I have been closely associated with the development of linguistic studies in this country and I have derived constant encouragement and co-operation from him and from all his students; I have not seen another scholar like him capable of building up a school of linguistics and a vital tradition of scholarship and dedication to knowledge. I am quite sure that this great project which is being undertaken by this University as a part of the work of the Linguistics Department under his Chairmanship will flourish under his magic direction and reach a worthy conclusion as a lasting memorial to South Indian scholarship in Linguistics. Pāṇini and Tolkappiyam are the two guiding lights in our linguistic firmament. May their spirit inform and enlighten our paths and inspire the new generation to keep that light aflame for all time to come!

6300

MAHAMAHOPADHYAYA
Dr. U. V. SWAMINATHAIYAR LIBRARY.
TIRUVANMIYUR. MADRAS-41.

64